RODIN: A Generative Model for Sculpting 3D Digital Avatars Using Diffusion

Abstract

This paper presents a 3D diffusion model that automatically generates 3D digital avatars represented as neural radiance fields (NeRFs). A significant challenge for 3D diffusion is that the memory and processing costs are prohibitive for producing high-quality results with rich details. To tackle this problem, we propose the roll-out diffusion network (RODIN), which takes a 3D NeRF model represented as multiple 2D feature maps and rolls out them onto a single 2D feature plane within which we perform 3D-aware diffusion. The RODIN model brings much-needed computational efficiency while preserving the integrity of 3D diffusion by using 3D-aware convolution that attends to projected features in the 2D plane according to their original relationships in 3D. We also use latent conditioning to orchestrate the feature generation with global coherence, leading to high-fidelity avatars and enabling semantic editing based on text prompts. Finally, we use hierarchical synthesis to further enhance details. The 3D avatars generated by our model compare favorably with those produced by existing techniques. We can generate highly detailed avatars with realistic hairstyles and facial hair. We also demonstrate 3D avatar generation from image or text, as well as text-guided editability.